

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 10/589,905
Source: IFWP
Date Processed by STIC: 8/28/06

ENTERED



IFWP

RAW SEQUENCE LISTING

DATE: 08/28/2006

PATENT APPLICATION: US/10/589,905

TIME: 10:32:34

Input Set : A:\701049_sequence_listing.txt

Output Set: N:\CRF4\08282006\J589905.raw

```

3 <110> APPLICANT: Yamanaka, Shinya
4      Dainippon Sumitomo Pharma Co., Ltd.
6 <120> TITLE OF INVENTION: Screening method for somatic cell nuclear reprogramming
substance
8 <130> FILE REFERENCE: 701049
C--> 10 <140> CURRENT APPLICATION NUMBER: US/10/589,905
C--> 10 <141> CURRENT FILING DATE: 2006-08-18
10 <150> PRIOR APPLICATION NUMBER: JP 2004-042337
11 <151> PRIOR FILING DATE: 2004-02-19
13 <150> PRIOR APPLICATION NUMBER: JP 2004-232961
14 <151> PRIOR FILING DATE: 2004-08-10
16 <150> PRIOR APPLICATION NUMBER: JP 2004-276572
17 <151> PRIOR FILING DATE: 2004-09-24
19 <160> NUMBER OF SEQ ID NOS: 50
21 <170> SOFTWARE: PatentIn Ver. 2.1
23 <210> SEQ ID NO: 1
24 <211> LENGTH: 1623
25 <212> TYPE: DNA
26 <213> ORGANISM: Mus musculus
28 <220> FEATURE:
29 <221> NAME/KEY: CDS
30 <222> LOCATION: (50)..(1369)
32 <400> SEQUENCE: 1
33 tgactgatct tgagtttgca taggcttcct gcggtgaaac gggtacact atg gcc tct 58
34                                     Met Ala Ser
35                                     1
37 ctg aag agg ttt cag acg ctc gtg ccc ctg gat cac aaa caa ggt acc 106
38 Leu Lys Arg Phe Gln Thr Leu Val Pro Leu Asp His Lys Gln Gly Thr
39      5      10      15
41 tta ttt gaa att att gga gag ccc aag ttg ccc aag tgg ttc cat gtc 154
42 Leu Phe Glu Ile Ile Gly Glu Pro Lys Leu Pro Lys Trp Phe His Val
43 20      25      30      35
45 gaa tgc ctg gaa gat cca aaa aga ctg tac gtg gaa cct cgg cta ctg 202
46 Glu Cys Leu Glu Asp Pro Lys Arg Leu Tyr Val Glu Pro Arg Leu Leu
47      40      45      50
49 gaa atc atg ttt ggt aag gat gga gag cac atc cca cat ctt gaa tct 250
50 Glu Ile Met Phe Gly Lys Asp Gly Glu His Ile Pro His Leu Glu Ser
51      55      60      65
53 atg ttg cac acc ctg ata cat gtg aac gtg tgg ggc cct gaa agg cga 298
54 Met Leu His Thr Leu Ile His Val Asn Val Trp Gly Pro Glu Arg Arg
55      70      75      80
57 gct gag att tgg ata ttc gga ccg ccg cct ttc cga agg gac gtt gac 346
58 Ala Glu Ile Trp Ile Phe Gly Pro Pro Pro Phe Arg Arg Asp Val Asp
59      85      90      95

```

RAW SEQUENCE LISTING

DATE: 08/28/2006

PATENT APPLICATION: US/10/589,905

TIME: 10:32:34

Input Set : A:\701049_sequence_listing.txt

Output Set: N:\CRF4\08282006\J589905.raw

```

61 cgg atg ctc act gat ctg gct cac tat tgc cgc atg aaa ctg atg gaa 394
62 Arg Met Leu Thr Asp Leu Ala His Tyr Cys Arg Met Lys Leu Met Glu
63 100 105 110 115
65 ata gag gct ctg gag gct gga gtt gag cgt cgt cgt atg gcg gcc cat 442
66 Ile Glu Ala Leu Glu Ala Gly Val Glu Arg Arg Met Ala Ala His
67 120 125 130
69 aag gct gcc acc cag cct gct ccc gtg aag gtc cgc gag gct gcc cct 490
70 Lys Ala Ala Thr Gln Pro Ala Pro Val Lys Val Arg Glu Ala Ala Pro
71 135 140 145
73 cgg ccc gct tcc gtg aag gtc cct gag acg gcc acc cag cct gct ccc 538
74 Arg Pro Ala Ser Val Lys Val Pro Glu Thr Ala Thr Gln Pro Ala Pro
75 150 155 160
77 gtg aag gtc cgc gag gct gcc cct cag ccc gct ccg gtg cag gag gtc 586
78 Val Lys Val Arg Glu Ala Ala Pro Gln Pro Ala Pro Val Gln Glu Val
79 165 170 175
81 cgc gag gct gcc cct cag cag gct tcc gtg cag gag gag gtc cgc gag 634
82 Arg Glu Ala Ala Pro Gln Gln Ala Ser Val Gln Glu Glu Val Arg Glu
83 180 185 190 195
85 gct gcc acc gag cag gct ccc gtg cag gag gtc cgc gag gct gcc acc 682
86 Ala Ala Thr Glu Gln Ala Pro Val Gln Glu Val Arg Glu Ala Ala Thr
87 200 205 210
89 gag cag gct ccc gtg cag gag gtc agc gag gct gcc acc gag cag gct 730
90 Glu Gln Ala Pro Val Gln Glu Val Ser Glu Ala Ala Thr Glu Gln Ala
91 215 220 225
93 ccc gtg cag gag gtc aac gag gct gcc acc gag cag gct tcc gtg cag 778
94 Pro Val Gln Glu Val Asn Glu Ala Ala Thr Glu Gln Ala Ser Val Gln
95 230 235 240
97 gcg gtc cgc gag gct gcc acc cgg ccg gct ccc ggg aag gtc cgc aag 826
98 Ala Val Arg Glu Ala Ala Thr Arg Pro Ala Pro Gly Lys Val Arg Lys
99 245 250 255
101 gcg gcc acc cag ccg gct ccg gtg cag gtt tgc cag gag gcc acc cag 874
102 Ala Ala Thr Gln Pro Ala Pro Val Gln Val Cys Gln Glu Ala Thr Gln
103 260 265 270 275
105 ttg gct ccc gtg aag gtc cgc gag gcg gcc acc cag ccg gct tcc ggg 922
106 Leu Ala Pro Val Lys Val Arg Glu Ala Ala Thr Gln Pro Ala Ser Gly
107 280 285 290
109 aag gtc cgc gag gcg gcc acc cag ttg gct cct gtg aag gtc cgc aag 970
110 Lys Val Arg Glu Ala Ala Thr Gln Leu Ala Pro Val Lys Val Arg Lys
111 295 300 305
113 gca gcc acc cag ttg gct cct gtg aag gtc cac gag gcg gcc acc cag 1018
114 Ala Ala Thr Gln Leu Ala Pro Val Lys Val His Glu Ala Ala Thr Gln
115 310 315 320
117 ccg gct ccg ggg aag gtc agc gat gct gcc acg cag tcg gct tcg gtg 1066
118 Pro Ala Pro Gly Lys Val Ser Asp Ala Ala Thr Gln Ser Ala Ser Val
119 325 330 335
121 cag gtt cgt gag gct gcc acg cag ctg tct ccc gtg gag gcc act gat 1114
122 Gln Val Arg Glu Ala Ala Thr Gln Leu Ser Pro Val Glu Ala Thr Asp
123 340 345 350 355
125 act agc cag ttg gct cag gtg aag gct gat gaa gcc ttt gcc cag cac 1162

```

RAW SEQUENCE LISTING

DATE: 08/28/2006

PATENT APPLICATION: US/10/589,905

TIME: 10:32:34

Input Set : A:\701049_sequence_listing.txt

Output Set: N:\CRF4\08282006\J589905.raw

```

126 Thr Ser Gln Leu Ala Gln Val Lys Ala Asp Glu Ala Phe Ala Gln His
127          360          365          370
129 act tca ggg gag gcc cac cag gtt gcc aat ggg cag tct ccc att gaa 1210
130 Thr Ser Gly Glu Ala His Gln Val Ala Asn Gly Gln Ser Pro Ile Glu
131          375          380          385
133 gtc tgt gag act gcc acc ggg cag cat tct cta gat gtc tct agg gcc 1258
134 Val Cys Glu Thr Ala Thr Gly Gln His Ser Leu Asp Val Ser Arg Ala
135          390          395          400
137 ttg tcc cag aag tgt cct gag gtt ttt gag tgg gag acc cag agt tgt 1306
138 Leu Ser Gln Lys Cys Pro Glu Val Phe Glu Trp Glu Thr Gln Ser Cys
139          405          410          415
141 ttg gat ggc agc tat gtc ata gtt cag cct cca agg gat gcc tgg gaa 1354
142 Leu Asp Gly Ser Tyr Val Ile Val Gln Pro Pro Arg Asp Ala Trp Glu
143 420          425          430          435
145 tca ttt atc ata tta taaatgcac tctggtgtga gccaggatag atggtacacg 1409
146 Ser Phe Ile Ile Leu
147          440
149 tctgcaaatac cagaacctaa aggcaggggt tagcttgggc tgagtaaggc aatgatctta 1469
151 aacctcagcc tgcctaagac tcccttcac tttctttctg gtttttggcc taggaatcgg 1529
153 gaagaacaga gtagagctgt ttttgtttcc ccattgtgtt aaatgtttgc agacacaatt 1589
155 taaagtattc taataaaaaa aaaattgcat tccc 1623
158 <210> SEQ ID NO: 2
159 <211> LENGTH: 440
160 <212> TYPE: PRT
161 <213> ORGANISM: Mus musculus
163 <400> SEQUENCE: 2
164 Met Ala Ser Leu Lys Arg Phe Gln Thr Leu Val Pro Leu Asp His Lys
165 1 5 10 15
167 Gln Gly Thr Leu Phe Glu Ile Ile Gly Glu Pro Lys Leu Pro Lys Trp
168 20 25 30
170 Phe His Val Glu Cys Leu Glu Asp Pro Lys Arg Leu Tyr Val Glu Pro
171 35 40 45
173 Arg Leu Leu Glu Ile Met Phe Gly Lys Asp Gly Glu His Ile Pro His
174 50 55 60
176 Leu Glu Ser Met Leu His Thr Leu Ile His Val Asn Val Trp Gly Pro
177 65 70 75 80
179 Glu Arg Arg Ala Glu Ile Trp Ile Phe Gly Pro Pro Pro Phe Arg Arg
180 85 90 95
182 Asp Val Asp Arg Met Leu Thr Asp Leu Ala His Tyr Cys Arg Met Lys
183 100 105 110
185 Leu Met Glu Ile Glu Ala Leu Glu Ala Gly Val Glu Arg Arg Met
186 115 120 125
188 Ala Ala His Lys Ala Ala Thr Gln Pro Ala Pro Val Lys Val Arg Glu
189 130 135 140
191 Ala Ala Pro Arg Pro Ala Ser Val Lys Val Pro Glu Thr Ala Thr Gln
192 145 150 155 160
194 Pro Ala Pro Val Lys Val Arg Glu Ala Ala Pro Gln Pro Ala Pro Val
195 165 170 175
197 Gln Glu Val Arg Glu Ala Ala Pro Gln Gln Ala Ser Val Gln Glu Glu

```

RAW SEQUENCE LISTING

DATE: 08/28/2006

PATENT APPLICATION: US/10/589,905

TIME: 10:32:34

Input Set : A:\701049_sequence_listing.txt

Output Set: N:\CRF4\08282006\J589905.raw

```

198          180          185          190
200 Val Arg Glu Ala Ala Thr Glu Gln Ala Pro Val Gln Glu Val Arg Glu
201          195          200          205
203 Ala Ala Thr Glu Gln Ala Pro Val Gln Glu Val Ser Glu Ala Ala Thr
204          210          215          220
206 Glu Gln Ala Pro Val Gln Glu Val Asn Glu Ala Ala Thr Glu Gln Ala
207 225          230          235          240
209 Ser Val Gln Ala Val Arg Glu Ala Ala Thr Arg Pro Ala Pro Gly Lys
210          245          250          255
212 Val Arg Lys Ala Ala Thr Gln Pro Ala Pro Val Gln Val Cys Gln Glu
213          260          265          270
215 Ala Thr Gln Leu Ala Pro Val Lys Val Arg Glu Ala Ala Thr Gln Pro
216          275          280          285
218 Ala Ser Gly Lys Val Arg Glu Ala Ala Thr Gln Leu Ala Pro Val Lys
219          290          295          300
221 Val Arg Lys Ala Ala Thr Gln Leu Ala Pro Val Lys Val His Glu Ala
222 305          310          315          320
224 Ala Thr Gln Pro Ala Pro Gly Lys Val Ser Asp Ala Ala Thr Gln Ser
225          325          330          335
227 Ala Ser Val Gln Val Arg Glu Ala Ala Thr Gln Leu Ser Pro Val Glu
228          340          345          350
230 Ala Thr Asp Thr Ser Gln Leu Ala Gln Val Lys Ala Asp Glu Ala Phe
231          355          360          365
233 Ala Gln His Thr Ser Gly Glu Ala His Gln Val Ala Asn Gly Gln Ser
234          370          375          380
236 Pro Ile Glu Val Cys Glu Thr Ala Thr Gly Gln His Ser Leu Asp Val
237 385          390          395          400
239 Ser Arg Ala Leu Ser Gln Lys Cys Pro Glu Val Phe Glu Trp Glu Thr
240          405          410          415
242 Gln Ser Cys Leu Asp Gly Ser Tyr Val Ile Val Gln Pro Pro Arg Asp
243          420          425          430
245 Ala Trp Glu Ser Phe Ile Ile Leu
246          435          440
250 <210> SEQ ID NO: 3
251 <211> LENGTH: 1063
252 <212> TYPE: DNA
253 <213> ORGANISM: Homo sapiens
255 <220> FEATURE:
256 <221> NAME/KEY: CDS
257 <222> LOCATION: (54)..(704)
259 <400> SEQUENCE: 3
260 tcggcctttg ggtttgctgt ggtgtccttg tctcctgcag gaccggccgc agc atg      56
261                                     Met
262                                     1
264 gac gct ccc agg cgg ttt ccg acg ctc gtg caa ctg atg cag cca aaa      104
265 Asp Ala Pro Arg Arg Phe Pro Thr Leu Val Gln Leu Met Gln Pro Lys
266          5          10          15
268 gca atg cca gtg gag gtg ctc ggt cac ctc cct aag cgg ttc tcc tgg      152
269 Ala Met Pro Val Glu Val Leu Gly His Leu Pro Lys Arg Phe Ser Trp

```

RAW SEQUENCE LISTING

DATE: 08/28/2006

PATENT APPLICATION: US/10/589,905

TIME: 10:32:34

Input Set : A:\701049_sequence_listing.txt

Output Set: N:\CRF4\08282006\J589905.raw

```

270          20          25          30
272 ttc cac tct gag ttc ctg aag aat ccg aag gta gtt cgc ctt gag gtt 200
273 Phe His Ser Glu Phe Leu Lys Asn Pro Lys Val Val Arg Leu Glu Val
274          35          40          45
276 tgg ctg gtg gaa aag atc ttc ggc cgg ggc gga gaa cgc atc ccg cac 248
277 Trp Leu Val Glu Lys Ile Phe Gly Arg Gly Glu Arg Ile Pro His
278 50          55          60          65
280 gtc cag ggt atg tcc caa atc ttg att cac gtg aat cga ttg gac cct 296
281 Val Gln Gly Met Ser Gln Ile Leu Ile His Val Asn Arg Leu Asp Pro
282          70          75          80
284 aac ggc gag gct gag atc ttg gta ttt ggg agg cct tct tac cag gag 344
285 Asn Gly Glu Ala Glu Ile Leu Val Phe Gly Arg Pro Ser Tyr Gln Glu
286          85          90          95
288 gac aca atc aag atg atc atg aac ctg gct gac tat cac cgc cag ctc 392
289 Asp Thr Ile Lys Met Ile Met Asn Leu Ala Asp Tyr His Arg Gln Leu
290          100          105          110
292 cag gcg aaa ggc tca gga aag gcc ctc gcc cag gat gtc gcc act cag 440
293 Gln Ala Lys Gly Ser Gly Lys Ala Leu Ala Gln Asp Val Ala Thr Gln
294          115          120          125
296 aag gcc gag acc cag cgg tct tca ata gaa gtc cgg gag gcc ggg acg 488
297 Lys Ala Glu Thr Gln Arg Ser Ser Ile Glu Val Arg Glu Ala Gly Thr
298 130          135          140          145
300 cag cgt tcg gtg gag gtc cgg gag gcc ggg acc cag cgt tcg gtg gaa 536
301 Gln Arg Ser Val Glu Val Arg Glu Ala Gly Thr Gln Arg Ser Val Glu
302          150          155          160
304 gtc cag gag gtc ggg aca cag ggt tct ccg gtg gag gtg cag gag gcc 584
305 Val Gln Glu Val Gly Thr Gln Gly Ser Pro Val Glu Val Gln Glu Ala
306          165          170          175
308 ggg acc cag cag tct ctc cag gct gcc aac aag tcg ggg acc cag cga 632
309 Gly Thr Gln Gln Ser Leu Gln Ala Ala Asn Lys Ser Gly Thr Gln Arg
310          180          185          190
312 tcc ccc gaa gct gcc agc aag gca gtg acc cag cgg ttt cgc gag gat 680
313 Ser Pro Glu Ala Ala Ser Lys Ala Val Thr Gln Arg Phe Arg Glu Asp
314          195          200          205
316 gcc cgg gac cca gtt act aga tta tgaaggcatc tcaggccctg gagccagagc 734
317 Ala Arg Asp Pro Val Thr Arg Leu
318 210          215
320 cagtcagggg ttaaagtga agcccgatt tccgccaga agctgggggt ggggagagga 794
322 tgtggatttt ttgttttacc ctttctgttg catggttgca aacacaaact tgagttctaa 854
324 taaagaattg caaagtggaa gcccgcccc cccctcccc ccgcctccct taagtccagg 914
326 aagctggggg ggcgaggaag gatgatgtg attgttttg tttaccct tttgttgat 974
328 ggttgccaac ccaaacttga gttttaataa ataattgcct ttccaaaaa aaaaaaaaaa 1034
330 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1063
333 <210> SEQ ID NO: 4
334 <211> LENGTH: 217
335 <212> TYPE: PRT
336 <213> ORGANISM: Homo sapiens
338 <400> SEQUENCE: 4
339 Met Asp Ala Pro Arg Arg Phe Pro Thr Leu Val Gln Leu Met Gln Pro

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/589,905

DATE: 08/28/2006

TIME: 10:32:35

Input Set : A:\701049_sequence_listing.txt

Output Set: N:\CRF4\08282006\J589905.raw

L:10 M:270 C: Current Application Number differs, Replaced Current Application No

L:10 M:271 C: Current Filing Date differs, Replaced Current Filing Date